Early Detection of Tuberculosis Using Machine Learning Techniques on Chest X-Ray Images

Dr.N.Umapathi
Professor
Dept. of Electronics and Communication Engineering
Jyothishmathi Institute of Technology and Science (JNTUH)
Karimnagar, Telanagana, India

G.Archana
UG Student
Dept. of Electronics and Communication Engineering
Jyothishmathi Institute of Technology and Science (JNTUH)
Karimnagar, Telanagana, India

B.Anand
UG Student
Dept. of Electronics and Communication Engineering
Jyothishmathi Institute of Technology and Science (JNTUH)
Karimnagar, Telanagana, India

V.Jahnavi
UG Student
Dept. of Electronics and Communication Engineering
Jyothishmathi Institute of Technology and Science (JNTUH)
Karimnagar, Telanagana, India

S.Shivateja
UG Student
Dept. of Electronics and Communication Engineering
Jyothishmathi Institute of Technology and Science (JNTUH)
Karimnagar, Telanagana, India

Abstract— **Tuberculosis (TB) continues to be one of the leading infectious diseases affecting millions of people globally, particularly in developing countries. Timely diagnosis plays a critical role in controlling transmission and reducing mortality. Chest X-ray imaging is widely used for pulmonary tuberculosis screening; however, manual interpretation depends heavily on expert radiologists and may lead to delays or subjective errors. In this paper, an automated tuberculosis detection system based on machine learning techniques is presented. The proposed model combines Convolutional Neural Networks (CNN) for deep feature extraction with Support Vector Machine (SVM) for classification. Preprocessing techniques such as resizing, normalization, and augmentation are applied to enhance dataset quality and improve generalization. The hybrid approach allows effective extraction of visual patterns associated with TB infection and produces reliable classification results. Experimental findings demonstrate that the system achieves high accuracy while maintaining computational efficiency. The proposed solution can serve as a supportive diagnostic tool in resource-limited healthcare environments**.

Index Terms-Tuberculosis Detection, Chest X-ray Analysis, Convolutional Neural Network, Support Vector Machine, Medical Image Classification

I. INTRODUCTION

Tuberculosis is a contagious bacterial disease primarily affecting the lungs. Despite advancements in medicine, TB remains a major public health concern worldwide. Early diagnosis is essential to prevent further spread and initiate timely treatment.

Chest radiography is one of the most common screening methods for pulmonary tuberculosis. However, interpreting radiographs requires specialized expertise and experience. In several rural and underdeveloped regions, trained radiologists may not be available, leading to delayed diagnoses and increased risk of complications.

The growth of artificial intelligence in medical imaging has opened new possibilities for automated disease detection. Machine learning models, especially deep learning networks, have demonstrated strong capabilities in recognizing patterns within image data. Convolutional Neural Networks can learn hierarchical image features automatically, reducing the need for manual feature engineering. Support Vector Machines provide robust classification by identifying optimal decision boundaries. The objective of this work is to design an automated system capable of identifying tuberculosis from chest X-ray images using a combination of CNN and SVM models. The system aims to improve diagnostic support, minimize dependency on manual screening, and assist healthcare professionals in early detection.

II. LITERATURE SURVEY

Lately theres been an effort to catch Tuberculosis (TB) by using medical imaging and machine learning. Researchers have been trying all sorts of machine learning. Deep learning tools on chest X-rays to spot TB infections with really good results.In the beginning most people used style machine learning methods for this job. They would clean up the X-ray images pull out the features and then let algorithms like Support Vector Machine (SVM) or K-Nearest Neighbors (KNN) do the sorting.They used things like texture analysis or edge detection—a lot of tricks—to spot anything in the lungs.These approaches worked okay. They relied a lot on manual feature engineering. That's not great especially since medical images can get pretty complicated. Then deep learning came along.Everything changed— with Convolutional Neural Networks (CNNs).Unlike old-style methods CNNs don't need people to pick features.They just learn what matters from the data.This shift made a difference.CNNs can dig deep picking up on patterns that older methods might miss. Thats pushed their accuracy way up in medical image classification.Studies using chest X-ray datasets have shown that these learning models do a great job at telling healthy lungs from ones with TB.Take one study for example.The researchers used CNNs. Transfer learning with big models like VGG16 ResNet50 and InceptionV3 to spot TB in chest X-rays.The results were around 90% accuracy sometimes even higher.That says a lot about how CNN-based methods have become good for this kind of work.Another team took things a step further.They built a custom CNN teamed it up with SVM working off the TBX11K dataset.They also used Principal Component Analysis (PCA) to pull out features before classifying the images.The SVM model hit 93.14% accuracy.The custom CNN topped it at 96.72%.That's a win for learning and TB detection.Some researchers have gotten more precise by using segmentation to focus on the lung area before running their models.By cropping out everything but the lungs they managed to push accuracy sometimes past 97%.Zeroing in like this helps the system ignore distractions and lock in on what matters for TB diagnosis.There's also an effort to make these AI tools transparent.Now researchers are combining machine learning with explainable AI so the models don't just flag TB—they actually highlight infected spots on the X-ray. This gives doctors a picture of why the model made its choice and helps build trust in the tech.Still it's not all easy.There are some challenges, like datasets, heavy computing power requirements and making sure models hold up when faced with new data.So researchers are determined to make TB detection systems smarter, faster and more reliable—. That work on TB detection is super important.

III. METHODOLOGY

Here's how it works. This system spots tuberculosis in chest X-rays by using both machine learning and deep learning. Everything starts with data—grabbing X-ray images from open medical databases. Some scans show healthy lungs, others have TB. Without this mix, nothing moves forward. The images get divided up: one set trains the model, the other tests how well it learned.Next, there's some image clean-up. Every X-ray gets resized to the same dimensions, and the pixel values are scaled so things line up. The system cuts down on noise and boosts contrast, which makes the details in the lungs stand out.Once the images are ready, the real work begins. The system hunts for features—textures, edges, anything that can help tell TB apart from healthy lungs. Regular machine learning goes after these patterns.Now, training kicks in. The system uses algorithms like Support Vector Machines (SVM) and deep learning models called Convolutional Neural Networks (CNNs). It studies the features, learns what sets TB apart, and just keeps improving as it sees more images.When it's trained up, the model faces new X-rays it hasn't seen before. Its job? Decide which ones are healthy and which ones have TB, based on what it's already learned.

Last step: see how it did. Metrics like accuracy, precision, recall, and F1-score show the results. If those numbers look solid, you've got a tool that helps doctors spot TB faster and more reliably..

VOLUME: 10 ISSUE: 03 | MARCH - 2026 SJIF RATING: 8.659 ISSN: 2582-3930

IV. PROPOSED SYSYEM

This system detects tuberculosis at an early stage by analyzing the images from the X-rays with the help of machine learning and deep learning techniques. The idea is quite simple: it helps doctors detect tuberculosis at an early stage with minimal effort. Instead of wasting a lot of time analyzing each and every X-ray, this system speeds up the process and makes things easier for the doctors.

Let's take a look at how it works. It all starts with the X-rays, which include images from both healthy and tuberculosis patients. The system processes these images with the help of a few steps: cleaning the images, extracting the required features, and identifying whether the images show signs of infection. The final output will be more precise with the help of these steps.

First and foremost, the system requires a good dataset. It uses X-rays from reliable sources and ensures that the dataset includes images from both healthy and tuberculosis patients. The system will be able to identify the difference, which is quite essential for the final output.

The images are the next step towards the final output. The system resizes the images and normalizes them so that they look similar. Then,

V. SYSTEM ARCHITECTURE

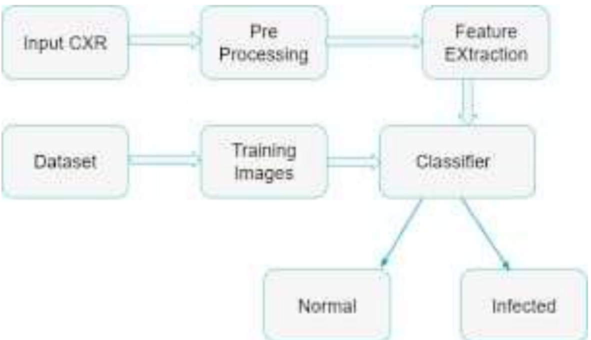


Fig:The diagram represents the complete workflow of a Tuberculosis (TB) detection system using Chest X-Ray (CXR) images and machine learning algorithms such as CNN and SVM

. The process begins with the input of a chest X-ray image, which may represent either a normal lung or a TB-infected lung. These images are collected from medical datasets or hospitals and may vary in size, format, brightness, and quality. Since raw images cannot be directly used for classification, they first undergo a preprocessing stage.

VI RESULT

Fig: Predicted ouput images

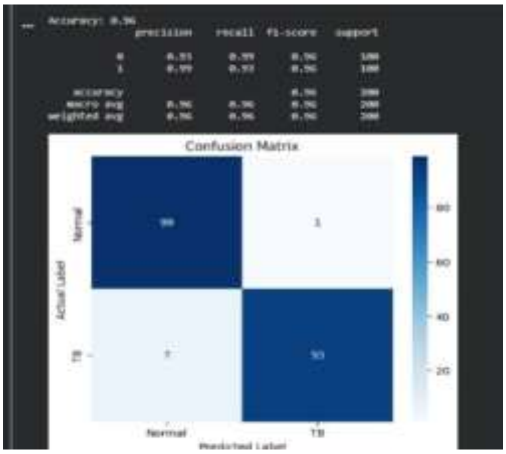
For each uploaded image, the system displays: The input chest X-ray image. The predicted class label (TB / Normal)

The prediction probability score. The prediction probability is calculated using the SVM decision function, which indicates how confidently the model classifies the image. A probability value close to 1 indicates strong confidence in the predicted class, while a value close to 0 indicates

Confusion Matrix:

shows the prediction distribution of a given chest X-ray image. The pie chart indicates that the Figure model predicted the image as Tuberculosis with a probability of 100%, while the probability for the Normal class is 0%. The system also displays the final result as "Prediction: Tuberculosis," confirming that the uploaded X-ray image is TB positive.The prediction confidence graph further illustrates that the probability value for Tuberculosis is

1.0, indicating very high confidence of the SVM model in its decision.



Prediction distribution and confidence matrix:

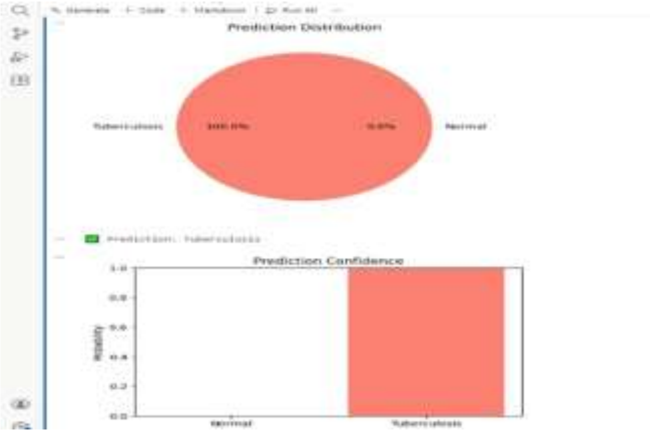


Fig:Confusion and pediction distribution & confidence matrix

VI. CONCLUSION AND FUTURE SCOPE

In this project, a machine learning-based system was developed for the automatic detection of Tuberculosis (TB) from chest X-ray images using the Support Vector Machine (SVM) algorithm. The model was trained using labeled X-ray images collected from the Kaggle dataset, consisting of TB-infected and normal cases. Proper preprocessing techniques such as image resizing, feature extraction, and normalization were applied to improve model performance. The dataset was divided into training and testing sets to evaluate the efficiency of the classifier. The trained SVM model successfully identified the optimal decision boundary to distinguish between infected and non-infected lungs. The system achieved an overall accuracy of 96%, with high precision, recall, and F1-score values for both classes. The confusion matrix analysis further confirmed that the number of misclassifications was minimal. The developed system demonstrates that machine learning techniques can effectively assist in early detection of Tuberculosis using chest X-ray images. This approach can serve as a supportive diagnostic tool for healthcare professionals, especially in areas where expert radiologists are limited.

Although the proposed system shows promising results, there are several opportunities for further improvement and enhancement. In the future, the system can be trained using a larger and more diverse dataset to improve generalization and robustness. Incorporating advanced deep learning models such as Convolutional Neural Networks (CNN) or transfer learning techniques may further increase classification accuracy. The system can also be extended to detect multiple lung diseases such as pneumonia, lung cancer, or COVID-19 from chest X-ray images. Integration of the model into a web-based or mobile application would allow real-time TB screening in rural and remote healthcare centers.

VII. REFERENCE

1.WorldHealthOrganization(2018).Globaltuberculosisreport2018.WorldHealth Organization. http://www.who.int/iris/handle/10665/274453.

2.Fatima,S.(2018).AutomatedTuberculosisDetectionandAnalysisUsingCXR's Images. Int. J. Comput. Electr. Eng.

3.Karnkawinpong, T., and Limpiyakorn,Y.(2019).Classificationofpulmonary tuberculosis lesion with convolutionalneuralnetworks.InJournalofPhysics: Conference Series.

4.Lakhani, P., and Sundaram, B. (2017). Deep learning at chest radiography: Automatedclassificationofpulmonarytuberculosis byusingconvolutional neural networks. Radiology 284, 574–582.

5.Qin, C.,Yao, D., Shi,Y., and Song, Z. (2018). Computer-aided detection in chest radiographybasedonartificialintelligence:Asurvey.Biomed.Eng.Online17,1– 23.

6.Sivaramakrishnan,R.,Antani,S.,Candemir,S.,Xue,Z.,Thoma,G.,Alderson,P., Abuya,J.,andKohli,M.(2018).Comparingdeeplearningmodelsforpopulation screening using chest radiography. In.

7."kaggle"(2019).NoTitle.Availableat: www.kaggle.com[AccessedOctober26, 2019].

8. Razavian,A.S.,Azizpour,H.,Sullivan,J.,andCarlsson,S.(2014).CNNfeatures off-the-shelf: An astounding baseline forrecognition.InIEEEComputerSociety Conference on Computer Vision and Pattern Recognition Workshops.

9. Liu, J., Liu,Y.,Wang, C., Li,A., Meng, B., Chai, X., and Zuo, P. (2018).An originalneuralnetworkforpulmonarytuberculosisdiagnosisinradiographs.In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).

10. Perera,P.,andPatel,V.M.(2019).LearningDeepFeaturesforOne-Class

11. Classification.IEEETrans.ImageProcess.

12. Hakim,B.,andBasari(2019).Tuberculosisdetectionanalysisusingtexture features on CXRs images. In AIP Conference Proceedings.

13. Krizhevsky,A.,Sutskever,I.,andHinton,G.E.(2012).ImageNetclassification with deep convolutional neural networks. InAdvances in Neural Information Processing Systems.

14. Perera,P.,andPatel,V.M.(2019).LearningDeepFeaturesforOne-Class Classification. IEEE Trans. Image Process.

15. Lu,D.,andWeng,Q.(2007).Asurveyofimageclassificationmethodsand techniques for improving classification performance. Int. J. Remote Sens.

16. Oquab,M.,Bottou,L.,Laptev,I.,andSivic,J.(2014). Learningandtransferring mid-level image representations using convolutional neural networks. In ProceedingsoftheIEEEComputerSocietyConferenceonComputerVisionand Pattern Recognition.